Bias Despite Masked Assessment of Clinical Outcomes When an Outcome Is Defined as One of Several Component Events

Sholom Wacholder, PhD, Jay H. Lubin, PhD, Mustafa Dosemeci, PhD, and Mitchell H. Gail, MD, PhD

National Cancer Institute, Epidemiology and Biostatistics Program, Maryland.

ABSTRACT: Hypothetical examples are presented that show that bias can result from misclassification of clinical outcomes defined as one of several events, even though the assessment of each component event was masked. In one example, the effect of the overall misclassification is to make a treatment that reduces the risk appear to increase it; in another example, misclassification causes the overall treatment effect to appear stronger than it actually is. These anomalies are due to the fact that the misclassification of the overall outcome can be differential, even though the misclassification of the individual components is nondifferential.

KEY WORDS: *Blinded study, diagnosis, misclassification, study design*

The purpose of masked assessment of disease outcome in clinical trials is the elimination of bias caused by conscious or subconscious factors in the assessment of outcomes [1]. In unmasked trials, disease may be overascertained or underascertained for one treatment but not the other ("differential misclassification"), resulting in the possibility of obtaining estimates of treatment effect which are, on average, more extreme or in the opposite direction from the true treatment effect. It is usually assumed that masking can prevent *differential* misclassification of disease and thereby prevent bias from exaggerating treatment effect or reversing the direction of treatment effects, since estimates of treatment effect can only be biased towards a lack of association under *nondifferential* misclassification.

In this article, we show that when a dichotomous disease outcome is defined as the presence of any of several mutually exclusive conditions, masking does not protect against bias from differential misclassification, because, even

---

Address reprint requests to: Sholom Wacholder, PhD, Epidemiology and Biostatistics Program, 6130 Executive Boulevard, EPN 403, Bethesda, MD 20892 USA.
Received September 18, 1990; revised February 13, 1991.

655 Avenue of the Americas, New York, New York 10010
0197-2456/91/$3.50

when misclassification is *nondifferential* for each component condition, it can be *differential* for the combined outcome. This phenomenon may result either in an exaggerated estimate of overall treatment effect or even a reversal in direction of the overall treatment effect. These findings have the same mathematical basis as our earlier work on the possible effects of exposure misclassification in epidemiologic studies [2]. For simplicity, we assume throughout this paper that component conditions may sometimes fail to be diagnosed (false-negative assessments) but that no persons without any condition will be mistakenly diagnosed as having a condition (no false-positive assessments); the same phenomena can occur when both false-negative and false-positive assessments are present.

We illustrate our point by discussing the results of a hypothetical clinical trial of whether a new prophylactic treatment A slows progression to AIDS in an HIV-positive group. In this example, AIDS is defined as the occurrence of either an opportunistic infection (OI) or other AIDS-defining conditions (OADC), but the argument could be extended to a detailed list of several AIDS defining conditions.

In the hypothetical trial, 1,000 patients were randomly assigned to treatment A, which was thought to protect against opportunistic infection, and another 1,000 to placebo. All were then followed for progression to AIDS. The numbers that would be obtained without misclassification are presented in Table 1a. The absolute *reduction* in AIDS due to treatment A is $0.95 - 0.90 = 0.05$; the risk of AIDS in the group treated with placebo is $0.95/0.90 = 1.056$ times higher than in group receiving treatment A. Note that treatment A reduces the risk of OI by $0.60 - 0.20 = 0.40$, but increases the risk of OADC by $0.70 - 0.35 = 0.35$, compared to the placebo-treated group. Now assume that the probability of misclassification to the category of no disease is 20% for both treatment groups, and both component outcomes, OI and OADC, as in Table 1b. The apparent effect of treatment A on OI is a reduction of only $0.48 - 0.16 = 0.32$, which is less than the 0.40 which would be found without misclassification. This is an example of the standard result [3] that nondifferential misclassification yields attenuation of the risk difference; since there is no misclassification from disease to nondisease in this example, there is no attenuation of the risk ratio, which remains at $3.0 = 0.48/0.16$. Now, suppose that in blind assessment of OI status, 20% of the OI cases in both groups are misclassified as nondiseased, but there is no misclassification of OADC, as in Table 1c. Note that the misclassification is *non*differential with respect to each component. If this were so, there would be reductions in the apparent benefit of treatment A with respect to OI and AIDS. Indeed, treatment A seems to be worse than placebo for AIDS: there is an apparent *excess* risk of $0.86 - 0.83 = 0.03$ and an unfavorable risk ratio of $0.83/0.86 = 0.965$ (Table 1c). The observed reversal occurred because the misclassification of disease is differential in the two-level AIDS versus no-AIDS categorization, even though it was nondifferential due to masking in the original three-level disease categorization. Specifically, the probability of misclassifying AIDS as no-AIDS is, from Table 1a $[(350/950) \times 0] + [600/950) \times 0.2] = 120/950 = 0.13$ for patients on placebo and $(700/900) \times 0 + (200/900) \times 0.2 = 40/900 = 0.04$ for those on treatment A.

Table 1 Hypothetical Effects of Nondifferential Misclassification of Disease After Collapsing the Disease Status

	Disease Status			
	None	OADC	OI	AIDS[a]
1a. True disease				
Number treated with placebo	50	350	600	950
Number treated with treatment *A*	100	700	200	900
Risk difference		−0.35	0.40	0.05
Risk ratio		0.5	3.0	1.056
1b. Equal Misclassification of OI and OADC[b]				
Number treated with placebo	240	280	480	760
Number treated with treatment *A*	280	560	160	720
Risk difference		−0.28	0.32	0.04
Risk ratio		0.5	3.0	1.056
1c. Misclassification of OI only[c]				
Number treated with placebo	170	350	480	830
Number treated with treatment *A*	140	700	160	860
Risk difference		−0.35	0.32	−0.03
Risk ratio		0.5	3.0	0.965
1d. Misclassification of OADC only[d]				
Number treated with placebo	120	280	600	880
Number treated with treatment *A*	240	560	200	760
Risk difference		−0.28	0.40	0.12
Risk ratio		0.5	3.0	1.158

[a]It is assumed that AIDS is diagnosed whenever one of the two mutually exclusive conditions, OADC or OI, is present.

[b]20% of patients with OI or OADC have been misclassified as nondiseased for each treatment. All other patients are classified correctly.

[c]20% of the patients with OI have been misclassified as nondiseased for each treatment. All other patients are classified correctly.

[d]20% of the patients with OADC have been misclassified as nondiseased for each treatment. All other patients are classified correctly.

Misclassification of components can also make the contrast between treatments more extreme. In Table 1d, 20% of patients with OADC have been misclassified as free of disease and all others are correctly diagnosed, giving another example of nondifferential misclassification with respect to each component condition. The apparent risk difference for AIDS between treatments is now increased to 0.88 − 0.76 = 0.12 instead of 0.05 and the observed relative risk is 1.156, instead of the less favorable value 1.056 in Table 1a.

We reported earlier [2] that grouping of exposure categories can result in spurious results in epidemiologic studies. Here we have shown that an analogous phenomenon can occur when disease categories are combined to make an overall outcome in clinical trials. The probability of misclassification of AIDS into no-AIDS is a weighted average of the probabilities of misclassification of OI and OADC into no-AIDS, where the weights will differ for the treated and placebo groups [2]. In general, whenever the distinct disease categories have different risks associated with treatment and different probabilities of misclassification into no-AIDS, differential misclassification will be induced (see the appendix of Ref. 2).

DISCUSSION

We have seen that misleading results can arise from diagnostic misclassification in a trial when classification of disease is based on two (or more) conditions, even though masking helps ensure that each condition is assessed equivalently in each treatment group. Sometimes these effects might be subtle or difficult to detect. For example, if a treatment is helpful for preventing recurrence of some histologic types of cancer and harmful for others, and if the probabilities of misclassifying these histologic types as noncancerous vary, anomalous findings may result, even though masked assessment ensures that the misclassification proportions are the same in each treatment group for each subtype.

In trials with nondifferential misclassification of the individual components, which masking is designed to achieve, bias away from the null or a reversal in direction of effect can only occur provided (1) each treatment is favored for at least one component condition and (2) the probabilities of misclassification vary across components. Misclassification into non-disease for a component whose treatment effect is in a direction *opposite* from the overall effect will usually attenuate the apparent effect of treatment on that component towards the null, thereby *enhancing* the magnitude of the overall effect. Misclassification into nondisease for a component whose treatment effect is in the *same* direction as the overall effect will typically attenuate the apparent effect of treatment on that component, thereby *diminishing* the overall effect; in extreme cases, as in Table 1c, the diminution can be extensive enough to actually change the direction of the apparent overall effect. If treatment acts favorably or unfavorably on all components, the result is an attenuated estimate of overall treatment effect, just as in the case of a single outcome category.

Our example illustrated a composite clinical outcome consisting of one or more of several clinical outcomes. A reviewer pointed out that these ideas might apply to other types of composite outcomes. For example, a composite outcome consisting of whether or not disease has progressed and whether or not toxicity has occurred might be used to assess drugs with serious toxicities, and a new drug might decrease the risk of disease progression and increase the risk of toxicity, compared to standard treatment. In this setting, nondifferential errors in measuring disease progression and toxicity might still result in anomalous findings.

In many applications, it may turn out that the phenomena we describe are of more theoretical than practical importance. Nonetheless, when an outcome is composed of several components, it seems prudent to examine treatment effects for each component separately as well as the overall disease outcome to determine whether treatment effects vary strikingly across the various components. Since there may be little power to detect such variations, particularly in the presence of misclassification, it is also especially important to minimize diagnostic error when a disease outcome is defined as any of several component conditions.

REFERENCES

1. Friedman LM, Furberg CD, DeMets DL: *Fundamentals of Clinical Trials*. Boston, John Wright PSG, 1981
2. Wacholder S, Dosemeci M, Lubin J. Blind exposure ascertainment does not guarantee nondifferential misclassification. *Am J Epidemiol*, in press
3. Newell DJ: Misclassification in 2 × 2 tables. *Biometrics* 19:187–188, 1963